title: "Divvy Cyclistic Bike Case Study"

author: "Barbara"

date: "2022-12-15"

# Introduction

This is a capstone project for the Google analytics Professional Certificate Program. This project prepares participants for a career in data analytics and focuses on analytics skills: data cleaning, analysis and visualization, along with the different tools to complete a business task. This project analyzes publicly available data on a bike sharing program in Chicago via divvy system website.

# Summary

In 2016, Cyclistic launched a successful bike sharing program. Since inception, the program has grown its fleet to 5,824 bicycles that are geo-tracked and locked into a network of 692 stations across Chicago. These bikes can be unlocked from one station and returned to any other station within the system at any time.

Cyclistic offers a flexible pricing plan from single, full day passes for casual riders or annual plan for members. A request for analysis on the types of riders and the difference between their bike usage and how marketing strategies can be used to convert casual riders to annual members was asked by the director of marketing.

# Ask Phase

Identify the business task: How do casual and annual members differ in their bike usage? Both casual riders and annual members are allowed to ride a bike within a day. However, casual riders subscribe to the pay as you go plan, where riders do not buy a plan, but pay for a single ride or day pass within a day as opposed to the annual members who make upfront fee for annual service to be render now or in the future.

Stakeholders Expectation: The stakeholders consist of cyclistic executive team, director of marketing -Lilly Moreno and cyclistic Analyst marketing team. Lilly Moreno asked for a marketing design strategy that aims at converting casual riders into annual membership as a key growth strategy.

# Prepare Phase

Divvy Bike Sharing system data is public and the source data can be found here: https://divvytripdata.s3.amazonaws.com/index.html. The data is organized in monthly, quarterly and yearly zip file which can be downloaded from the site and saved as excel or csv file. The data on the site is reliable and secure but incomplete, where some data values must be changed or derived. The data privacy issue prohibits the use of personal identifiable information, this prevents determining a rider PII related information and determining the number of past or present purchase made by a rider.

# Process Phase

The process of collecting, cleaning and examining the data was primarily done in R studio utilizing R programming with complementary cleaning done in Microsoft excel. To get familiar with data, a check was performed on the data structure, data types of definition, missing values, duplicates, data integrity and quality. The fictional data for this project was provided by Motivate International, Inc. as a public repository from which data was retrieved for this project.

# Data Dictionary

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Description | Datatype | Accept Nulls? |  |
| member\_casual | member or subscriber | char | no |  |
| rider\_id | number which identifies a ride | char |  |  |
| rideable\_types | electronic bike  classic \_bike  docked bike | char | no |  |
| start\_station\_id | A number or string that identifies the station at which a ride started | varchar | yes |  |
| start\_station\_name | The station at which the ride started |  |  |  |
| started\_at | Start date and time at which a ride started | date-timestamp | no |  |
| end\_station\_id | A number or string that identifies the station at which a ride ended | varchar | yes |  |
| end\_station\_name | Station name at which a ride ended | char | no |  |
| ended at | End date and time at which a ride ended | date-timestamp | no |  |
| start\_lon | Longitude coordinate at which a ride started | numeric | no |  |
| end\_lon | Longitude coordinate at which a ride ended | numeric | no |  |
| stat\_lat | Latitude coordinate at which a ride started | numeric | no |  |
| end\_lat | Latitude coordinate at which a ride started | numeric | no |  |

# Visualization

A graphical representation of the divvy trip is created to convey meaningful insight from the analytically findings on the differences between causal riders and annual members using primary pie chart and bar graph to show the differences between causal riders and annual members.

# Analysis

Casual members predominately ride on the weekend whereas annual members use different rides on weekdays.

Casual member ride duration is almost half the time as annual members.

Both casual rider and annual member utilize classic bikes the most, accompanied by electronic bike and docked bike as the least use of all bikes.

# Recommendations

Since casual rider ride time is almost half the time as annual members, a cost benefit analysis should be created to show causal rider how much money would be saved over time if they convert to annual membership.

* Create optional memberships, annual membership (356 days) or weekend only annual membership (53 weeks) to generate reoccurring revenue to maintain constant and consistent stream of revenue.
* Create a discount feature associated with annual memberships linked to festive periods, sports events, or concerts, etc.to give members easy access to these events within their area.
* Create seamless features that accumulate points and automatically allocate points to members to be used through the year or allow members to share the points with friends.
* Create features that keep members informed, educate them about bike usage, or related events, features which reward members for good behavior to enforce change in behaviors.